动 物 学 研 究 2007，Jun. 28（3）：271－278 CN 53－1040/Q ISSN 0254－5853
Zoological Research

miR-34 基因家族的分子进化

高佳莉，罗玉萍*，李思光
（南昌大学 生命科学学院，江西 南昌 330031）

摘要：根据 miRNA 基因在进化中高度保守的特点，利用生物信息学方法在目前已测序的动物物种中搜寻参与哺乳动物早期发育调控的 mir-34 基因的同源序列，在 33 个不同的动物物种中获得了 miR-34 基因的 54 条同源序列，其中 18 条为新发现的序列。表明 miR-34 是高度保守的，广泛存在于后生动物中。目前发现的 *mir-34* 基因 80%位于基因间隔区，少数位于蛋白编码基因的内含子区和 3′UTR 上。不同动物中，*mir-34* 基因成熟序列的同源性为 68%，前体序列为 38.89%。在无脊椎动物中只有一个 *mir-34*，而在几乎所有的脊椎动物中都有 *mir-34a*，*mir-34b*，*mir-34c*，形成 miR-34 基因家族。系统进化分析表明，脊椎动物中 miR-34 基因家族是通过基因的串联和局部重复形成的，这个过程中伴随着个别碱基的变异。

关键词：microRNA；miR-34；分子进化
中图分类号：Q523；Q588 **文章标识码**：A **文章编号**：0254-5853（2007）03-0271-08

Molecular Evolution of miR-34 Gene Family

GAO Jia-li，LUO Yu-ping*，LI Si-guang
（*College of Life Sciences*，*Nanchang University*，*Nanchang* 330031，*China*）

Abstract：miR-34 gene can regulate early development of mammalian. Using bioinformatics analysis of sequenced metazoan genomes，54 miR-34 genes of 33 different animal species were found. Eighteen miR-34 genes were identified in the first time，suggesting miRNA are highly conservative and widespread in Eukaryotes. Of these *mir-34* gene homologues identified in metazoan，80% locate in the intergenic region，the others are in the intron or 3′UTR of protein-coding gene. miR-34 gene is conserved in various animals，and the mature sequences and precursors of *mir-34* gene in metazoan analyzed share 68% and 38.89% sequence identity，respectively. However there is only a *mir-34* gene in invertebrate and three *mir-34* genes，*mir-34a*，*mir-34b* and *mir-34c* are present in almost all vertebrate. The phylogenetic analysis reveals that the miR-34 gene family of vertebrate originated from local and tandem duplications followed by mutation of individual base after divergence from invertebrate.

Key words：microRNA；miR-34；Molecular evolution

MicroRNA（miRNA）是一类由内源基因编码的长度约为 22 个核苷酸的非编码单链 RNA 分子，它们在动植物中参与转录后基因表达调控。第一个 miRNA 基因——*lin-4* 是 Lee et al（1993）在研究线虫发育缺陷时发现的，它通过与靶基因 *lin-14* mRNA3′UTR 区互补配对，从而抑制 Lin-14 蛋白质的表达。Reinhart et al（2000）和 Pasquinelli et al（2000）又在线虫中发现了第二个有异时性开关功能的 let-7 RNA，其靶基因控制着线虫从 L4 阶段到成虫阶段的发育转变。正像 lin-4 和 let-7 那样，动物 miRNA 主要是通过翻译抑制来负调控靶基因。不同的是，植物 miRNA 主要是引起靶基因 mRNA 的降解。

到目前为止，在动植物以及病毒中已经发现有 4361 个 miRNA 分子（Release 9.0，October 2006）。大多数 miRNA 基因以单拷贝、多拷贝或基因簇（cluster）的形式存在于基因组中（Lagos-Quintana et al，2001；Lau et al，2001）。然而，最近的研究表明大约 70%的哺乳动物 miRNA 基因是位于 TUs 区（transcription units，TUs）（Rodriguez et al，

* 收稿日期：2006-12-18；接受日期：2007-03-23
基金项目：国家自然科学基金资助项目（30660042）；江西省自然科学基金资助项目（0630136）
* 通讯作者（Corresponding author），Tel：0791-8304938，E-mail：luoyuping@163.com

2004），且其中大部分是位于内含子区（Kim & Nam，2006）。一些内含子 miRNA 基因的位置在不同的物种中是高度保守的，例如，miR-7 在哺乳动物和昆虫的 hnRNP K 蛋白内含子区都有发现。另外一个很有趣的例子是，miR-106b 基因家族在人类和老鼠中均位于 MCM7 基因的第 13 个内含子上（Kim & Nam，2006）。miRNA 不仅在基因位置上保守，序列上也呈现出高度的同源性（Pasquinelli et al，2000；Ruvkun et al，2001；Lee & Ambros，2001）。在线虫 *Caenorhaboditis elegans* 中所发现的 miRNA，85%都可以在 *C. briggsae* 的基因组中找到同源序列（Lau et al，2001）。植物中，miR-39 的同源序列在拟南芥、水稻和烟草中都存在（Reinhart et al，2002；Llave et al，2002）。

miRNA 高度的保守性与其功能的重要性有着密切的关系。Houbaviy et al（2003）研究发现，miR-34 基因在小鼠胚胎干细胞中特异性表达，它们可能参与维持胚胎细胞的全能性以及哺乳动物的早期发育调控。miRNA 与其靶基因的进化有着密切的联系，研究其进化历史有助于进一步了解其作用机制和功能。本文依据 miRNA 基因在物种间的保守性，利用生物信息学同源性搜寻、比对的方法分析 miR-34 基因家族的系统进化历史。

1 材料和方法

1.1 数据的获得

从 miRNA 数据库（http://microrna.sanger.ac.uk/cgi-bin/sequences/browse.pl）中获得人 *mir-34* 基因的前体序列（*has-mir-34a*，*has-mir-34b*，*has-mir-34c*）以及线虫 *mir-34* 基因的前体序列（*cel-mir-34*）。

1.2 同源性搜寻

分别用人和线虫的 miR-34 基因的前体序列通过同源性比对的方法在所有动物物种中进行搜寻。本文使用了三个数据库，分别是 NCBI 数据库（http://www.ncbi.nlm.nih.gov/blast）；ENSEMBL 数据库（http://www.ensembl.org/index.html）以及 UCSC 数据库（http://genome.ucsc.edu/cgi-bin/hgBlat）。在前两者中是通过 BlastN（Altschul et al，1990）程序进行搜寻，在 UCSC 数据库中是通过 BLAT（Kent，2002）程序搜寻。

1.3 二级结构预测

利用在线的 Mfold 程序（http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi）（Zuker，2003）对所有搜寻到的新序列进行二级结构的预测。将预测结果与 miRNA 的鉴定标准（Ambros et al，2003）作比较，成熟序列长度应约为 22 核苷酸，计算机模拟可以得到 miRNA 前体保守的二级结构，并且成熟序列应位于发夹结构茎区的 5′或 3′侧。通过序列的筛选，最终确定 miR-34 基因的同源序列。

1.4 同源性比对

采用 Clustal W 程序（Thompson et al，1994）对 miR-34 基因的所有同源序列进行多序列比对（multiple sequence alignment，MSA）。

1.5 构建系统发育树

最后，用 MEGA（3.1 版本）软件对搜寻到的 miR-34 基因及其同源序列进行多序列比对，并做出相应的系统进化树。

2 结果分析

2.1 miR-34 基因家族新成员的鉴定及特征分析

通过同源性搜索的方法，分别在 33 个不同的动物物种中，找到了 54 条 Mir-34 基因的同源序列（Tab. 1），其中 18 条为新发现的序列（斜体），它们都具有典型的茎环状二级结构（Fig. 1）。本文中一致用 miR-#代表 miRNA，*mir-#*（斜体）代表 miRNA 前体基因。miR-34 基因及其同系物广泛存在于节肢动物，线虫纲动物以及许多哺乳动物中，保守性很高。miR-34 家族中大部分成员（70%）的基因位置已知，其中 84%都位于基因间隔区（intergenic region，IGR），少数位于蛋白编码基因的内含子区和 3′UTR 上。狗 *mir-34a* 基因位于其 5 号染色体 LOC608460 基因的内含子区，原鸡 *mir-34a* 基因位于 ENSGALT00000035086/87 基因的内含子区。此外，人和黑猩猩的 *mir-34b* 和 *mir-34c* 基因分别位于同源基因的内含子区和 3′UTR 上。

序列分析发现在人、黑猩猩、牛、狗、小鼠、大鼠、斑马鱼以及原鸡 8 个物种中，*mir-34b* 基因与 *mir-34c* 基因均紧密相连，位于同一染色体上。人和黑猩猩的 *mir-34b* 和 *mir-34c* 基因分别位于同源基因的内含子区和 3′UTR 上，而在其他物种中 *mir-34b* 基因和 *mir-34c* 基因都位于同一基因间隔区内（Fig. 2）。这一组织形式的特征暗示了这两个基因（*mir-34b* 与 *mir-34c*）很可能是在进化过程中由串联重复产生的（Tanzer & Stadler，2004）。

2.2 miR-34 基因家族同源性分析

表 1 所有已测序的动物物种中的 miR-34 基因及其同源物

Tab. 1 miR-34 and its orthologue found in sequenced metazoan

miR 名称 miR name	物种 Species	基因位置 Gene location	基因组织形式 Gene organization	正/负链 +/− chanin
aga-miR-34 ▲☆	冈比亚按蚊 *Anopheles gambiae*	ch2R：28232717–28232737	基因间隔区	−
age-miR-34a●	蜘蛛猴 *Ateles geoffroyi*	未知	未知	未知
ame-miR-34 ▲☆	蜜蜂 *Apis mellifera*	group5：327549–327569	基因间隔区	+
bmo-miR-34▲	家蝉 *Bombyx mori*	BAAB01147222：1416–1513	基因间隔区	+
bta-miR-34a ▲☆	牛 *Bos taurus*	chr16：30629167–30629189	基因间隔区	−
bta-miR-34b▲☆	牛 *B. taurus*	Un. 7588：19678–19699	基因间隔区	−
bta-miR-34c▲☆	牛 *B. taurus*	Un. 7588：18978–18999	基因间隔区	−
cbr-miR-34▲☆	线虫 *Caenorhabditis briggsae*	chrUn：55197686–55197707	基因间隔区	+
cel-miR-34▲☆	线虫 *C. elegans*	chrx：2969803–2969824	基因间隔区	−
cfa-miR-34a ▲☆	狗 *Canis familiaris*	chr5：65484449–65484471	LOC608460 内含子区	−
cfa-miR-34b ▲☆	狗 *C. familiaris*	chr5：24566992–24567014	基因间隔区	−
cfa-miR-34c ▲☆	狗 *C. familiaris*	chr5：24567022–24567044	基因间隔区	−
dre-miR-34▲☆	斑马鱼 *Danio rerio*	ch23：24608106–24608127	基因间隔区	+
dre-miR-34b▲☆	斑马鱼 *D. rerio*	ch5：64601186–64601208	基因间隔区	−
dre-miR-34c▲☆	斑马鱼 *D. rerio*	ch5：64600946–64600968	基因间隔区	−
dan-miR-34 ☆	嗜凤梨果蝇 *Drosophila ananassae*	scaffold_13340：6488582–6488602	未知	+
der-miR-34 ☆	果蝇 *D. erecta*	scaffold_4770：15677094–15677114	未知	−
dgr-miR-34 ☆	果蝇 *D. grimshawi*	scaffold_24940：2264380–2264400	未知	+
dme-miR-34▲☆	果蝇 *D. melanogaster*	chr3R：5926674–5926695	基因间隔区	+
dpe-miR-34 ☆	果蝇 *D. persimilis*	super_19：725072–725092	未知	−
dps-miR-34▲☆	果蝇 *D. pseudoobscura*	chr2：4980658–4980679	基因间隔区	−
dse-miR-34 ☆	果蝇 *D. sechellia*	super_0：15986909–15986929	未知	−
dsi-miR-34 ☆	拟果蝇 *D. simulans*	ch3R：15449447–15449467	未知	−
dvi-miR-34 ☆	黑果蝇 *D. virilis*	scaffold_12855：2681947–2681967	未知	−
dya-miR-34 ☆	果蝇 *D. yakuba*	ch3R：9974486–9974506	未知	+
ete-miR-34a ●	无尾猬 *Echinops telfairi*	scaffold_39512：1412–1434	基因间隔区	+
gga-miR-34a▲☆	原鸡 *Gallus gallus*	ch21：3118953–3119075	ENSGALT00000035086/87 内含子区	−
gga-miR-34b▲☆	原鸡 *G. gallus*	ch24：5217102–5217124	基因间隔区	+
gga-miR-34c▲☆	原鸡 *G. gallus*	ch24：5217839–5217861	基因间隔区	+
ggo-miR-34a●	大猩猩 *Gorilla gorilla*	未知	未知	未知
hsa-miR-34a▲☆	人 *Homo sapiens*	ch1：9134380–9134402	基因间隔区	−
hsa-miR-34b▲☆	人 *H. sapiens*	ch11：110888885–110888907	LOC728196 内含子区	+
hsa-miR-34c▲☆	人 *H. sapiens*	ch11：110889386–110889408	3′UTR 区	+
ptr-miR-34a▲●	黑猩猩 *Pan troglodytes*	ch1：9151501–9151522	基因间隔区	−
ptr-miR-34b ▲●	黑猩猩 *P. troglodytes*	ch11：6998680–6998702	LOC738013 内含子区	+
ptr-miR-34c ▲●	黑猩猩 *P. troglodytes*	ch11：6999193–6999215	3′UTR 区	+
lla-miR-34a●	兔猴 *Lagothrix lagotricha*	未知	未知	未知
mdo-miR-34a▲	南美负鼠 *Monodelphis domestica*	ch4：428562614–428562724	基因间隔区	+
mml-miR-34a ▲	猕猴 *Macaca mulatta*	scaffold91438：18223–18244	基因间隔区	+
mne-miR-34a●	豚尾猴 *M. nemestrina*	未知	未知	未知
ppa-miR-34a●	矮黑猩猩 *Pan paniscus*	未知	未知	未知
ppy-miR-34a●	猩猩 *Pongo pygmaeus*	未知	未知	未知
mmu-miR-34a▲☆	小鼠 *Mus musculus*	ch4：148912273–148912295	基因间隔区	+
mmu-miR-34b▲☆	小鼠 *M. musculus*	ch9：50855841–50855863	基因间隔区	−
mmu-miR-34c▲☆	小鼠 *M. musculus*	ch9：50855306–50855328	基因间隔区	−
rno-miR-34a▲☆	大鼠 *Rattus norvegicus*	ch5：167187916–167187938	基因间隔区	+
rno-miR-34b▲☆	大鼠 *R. norvegicus*	ch8：54422619–54422641	基因间隔区	−
rno-miR-34c▲☆	大鼠 *R. norvegicus*	ch8：54422094–54422116	基因间隔区	−
sla-miR-34a●	红腹狨 *Saguinus labiatus*	未知	未知	未知
xtr-miR-34a▲☆	非洲爪蟾 *Xenopus tropicalis*	scaffold_207：56399–56421	未知	+
xtr-miR-34b-1▲☆	非洲爪蟾 *X. tropicalis*	scaffold_1127：6080–6102	基因间隔区	−
xtr-miR-34b-2▲☆	非洲爪蟾 *X. tropicalis*	scaffold_1127：13647–13669	基因间隔区	+
xtr-miR-34b-3▲☆	非洲爪蟾 *X. tropicalis*	scaffold_39：3202291–3202313	基因间隔区	−
xtr-miR-34b-4▲☆	非洲爪蟾 *X. tropicalis*	scaffold_39：3208151–3208173	基因间隔区	+

▲表示使用 NCBI 数据库搜寻获得，☆表示使用 UCSC 数据库，●表示使用 ENSEMBL 数据库。黑斜体表示新发现的 miR-34 同源序列。
Triangles denote data from NCBI. Asterisks denote data from UCSC. Rounds denote data from ENSEMBL. New orthologues of miR-34 are in black italics.

图 1 新 Mir-34 基因家族成员的二级结构预测

Fig. 1 Predicted hairpin secondary structures of identified novel miRNAs using RNAviz software

成熟 microRNA 序列用下划线表示（Mature miRNA is underlined）。

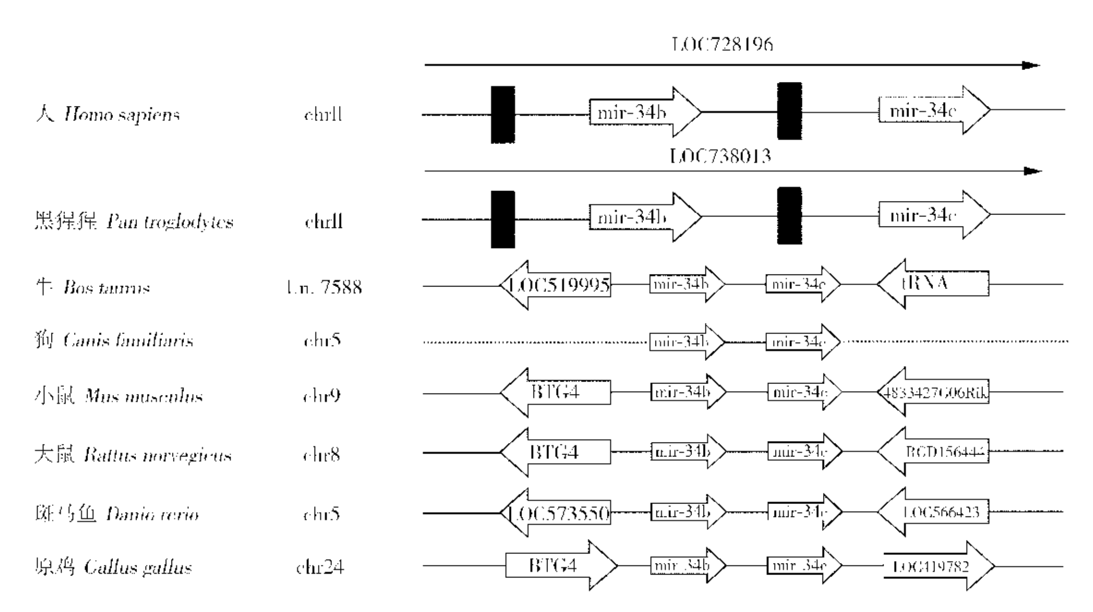


图 2 八个物种中 *mir-34b* 和 *mir-34c* 基因位置示意图

Fig. 2 Characterization of genomic organization of *mir-34b* and *mir-34c* in the eight animal species

方框表示外显子，实线箭头表示转录方向，空心箭头分别表示 *mir-34b*、*mir-34c* 基因及其上下游基因，虚线部分表示未知序列。

Exons are represented by boxes. *mir-34b*, *mir-34c* and neighbor gene sequences are represented by arrow. The cartoon is not drawn to scale.

miR-34 家族成员的成熟序列同源性很高，约为 68%（Tab. 2）。序列长度是相近的，22—23 nt，保守的碱基数为 19 个。miRNA 的 5′端的 2—8 个核苷酸对于 miRNA 与靶序列 3′ UTR 区的配对非常重要，因而被称为种子序列（Doench & Sharp，2004）。这段序列在同源的 miRNA 中通常是保守的，比对的结果正证明了这一点。从差异碱基的变化特征来看，miR-34 基因及其同系物在进化过程中很可能发生了碱基的缺失。例如 miR-34a 很可能在 11 位缺失了一个碱基。此外，10—11 位碱基在四个家族成员中保守性相对不高，这可能是由于两个碱基错配的结构更有利于 miRNA 与其靶序列的配对（Doench & Sharp，2004）。

多序列比对分析显示 miR-34 基因家族成员前体基因同源性约为 38.89%（Fig. 3）。图中星号部分表示了同源性较高的碱基。有少数位点的碱基呈现出类似于单核苷酸多态性（SNP）的现象（Tong et al，2006）。

表 2　MiR-34 基因家族各成员成熟序列比较

Tab. 2　Multiple sequence alignment of the mature sequence of miR-34 gene family

MIR 名称 MIR name	成熟序列 Mature sequence	碱基数（nt） Base number
miR-34	aggcagugugguuagcugguug	22
	uggcagugugguuagcugguug	
miR-34a	uggcaguguc-uuagcugguugu	23
miR-34b	uaggcagugucauuagcugauug	23
miR-34c	aggcaguguaguuagcugauugc	23

保守碱基用阴影标注，碱基的缺失用短线（-）表示。

Highly conserved bases are shadowed. Hyphen denotes base deletion.

```
cfa-34b     ---------TTTTTCTGTGAGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
hsa-34c     -------------------AGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
ptr-34c     -------------------AGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
bta-34c     -------------------AGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
mmu-34c     -------------------AGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
rno-34c     -------------------AGTCTAGTTACTAGGCAGTGTAGTTAGCTGATTGCT-----
gga-34c     -------------------AGCCTGGTTACCAGGCAGTGTAGTTAGCTGATTGCC-----
dre-34c     ----------------TGCTGTGTGGTCACCAGGCAGTGCAGTTAGTTGATTACA-----
xtr-34b-1   -------------------TGTTGGGTTTTCAGGCAGTGTAGTTAGCTGATTGTGTTAAC
xtr-34b-3   -------------------TGTTGGGTTTTCAGGCAGTGTAGTTAGCTGATTGTGTTAAC
xtr-34b-2   ----------------------------TTCAGGCAGTGTAGTTAGCTGATTGTGTTA-T
xtr-34b-4   ----------------------------TTCAGGCAGTGTAGTTAGCTGATTGTGTTA-T
dre-34b     ------------------GGGGTTGGTCTGTAGGCAGTGTTGTTAGCTGATTGTTT---C
bta-34a     ----------GGCCAGCTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
cfa-34a     TCATCGTCAGGGCCAGCTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
hsa-34a     ----------GGCCAGCTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
ptr-34a     ----------GGCCAGCTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
gga-34a     -----------GCCAGCTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
mmu-34a     ------------CCAGCTGTGAGTAATTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
rno-34a     ------------CCGGCTGTGAGTAATTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGAG
xtr-34a     ----------------CTGTGAGTGTTTCTTTGGCAGTGTC-TTAGCTGGTTGTTGTGGC
dre-34      -------------CTGCTGTGAGTGGTTCTCTGGCAGTGTC-TTAGCTGGTTGTTGTGTG
hsa-34b     ------------------GTGCTCGGTTTGTAGGCAGTGTCATTAGCTGATTGTACTGTG
ptr-34b     ------------------GTGCTCGGTTTGTAGGCAGTGTCATTAGCTGATTGTACTGTG
mmu-34b     ------------------GTGCTCGGTTTGTAGGCAGTGTAATTAGCTGATTGTAGTGCG
rno-34b     ------------------GTGCTCGGTTTGTAGGCAGTGTAATTAGCTGATTGTAGTGCG
bta-34b     ------------------GTGCTCGGTTTGTAGGCAGTGTAATTAGCTGATTGTACTCTC
cfa-34c     --------CGGGCGCCCGGTGCTCGGTTTGTAGGCAGTGTAATTAGCTGATTGTCCTCTG
gga-34b     ------------------GTGCTTGGTTTGCAGGCAGTGTAGTTAGCTGATTGTACCCAG
cbr-34      -------------AAGCACTCATGGTCGTGA-GGCAGTGTGGTTAGCTGGTTGCATACAC
cel-34      --------------CGGACA-ATGCTCGAGA-GGCAGTGTGGTTAGCTGGTTGCATATTT
dme-34      -------------AATT--GGCTATGCGCTTTGGCAGTGTGGTTAGCTGGTTGTGTAGCC
dps-34      -------------AATT--GGCTATGCGCTTTGGCAGTGTGGTTAGCTGGTTGTGTAGCC
bmo-34      -----------AGAATCAGGGTAGACCGCGTTGGCAGTGTGGTTAGCTGGTTGTGTATGG
                                            *******    **** ** **
```

```
cfa-34b     AATAGTAC-------CAATCACTAACCACACGGCCAGGTAAAAAGATTTGGGAATCAT--
hsa-34c     AATAGTAC-------CAATCACTAACCACACGGCCAGGTAAAAAGATT------------
ptr-34c     AATAGTAC-------CAATCACTAACCACACGGCCAGGTAAAAAGATT------------
bta-34c     AATAATAC-------CAATCACTAACCACACGGCCAGGTAAAAAGATT------------
mmu-34c     AATAGTAC-------CAATCACTAACCACACAGCCAGGTAAAAAGACT------------
rno-34c     AATAGTAC-------CAATCACTAACCACACAGCCAGGTAAAAAGACT------------
gga-34c     ACCAGGAC-------CAATCACTAACCACACAGCCAGGTAAAAAG---------------
dre-34c     ATCCATAAA-----GTAATCACTAACCTCACTACCAGGTGAAGGCTAGTA----------
xtr-34b-1   ATAAGACTT-----GCAATCACTAGCTAAACTACCAGCAAAACTAAACA-----------
xtr-34b-3   ATAAGACTT-----GCAATCACTAGCTAAACTACCAGCAAAACTAAACA-----------
xtr-34b-2   ATCAAATTT-----GCAATCACTAGCTAAACTACCATAAAA-------------------
xtr-34b-4   ATCAAATTT-----GCAATCACTAGCTAAACTACCATAAAA-------------------
dre-34b     ATATGAACT-----ATAATCACTAACCATACTGCCAACACAACAACCTACA---------
bta-34a     TAATAATGCAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACGTTGT-GGGC
cfa-34a     TAATAGTGAAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACGTTGTTGGGC
hsa-34a     CAATAGT-AAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACGTTGTGGGGC
ptr-34a     CAATAGT-AAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACGTTGTGGGGC
gga-34a     CAATAGTTAAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTACACATTGTTGGGC
mmu-34a     TATTAGCTAAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACATTGT-----
rno-34a     TATTAGCTAAGGAAGCAATCAGCAAGTATACTGCCCTAGAAGTGCTGCACGTTGT-----
xtr-34a     ACGTTATAGAAGTAGCAATCAGCAAATATACTGCCCTAGAAGTTCTGCACATT-------
dre-34      GAGTGAG-AACGAAGCAATCAGCAAGTATACTGCCGCAGAAACTCGTCACCTT-------
hsa-34b     GTGGTTA--------CAATCACTAACTCCACTGCCATCAAAACAAGGCAC----------
ptr-34b     GTGGTTA--------CAATCACTAACTCCACTGCCATCAAAACAAGGCAC----------
mmu-34b     GTGCTGA--------CAATCACTAACTCCACTGCCATCAAAACAAGGCAC----------
rno-34b     GTGCTCA--------CAATCACTAACTCCACTGCCATCAAAACAAGGCAC----------
bta-34b     ATGCTTA--------CAATCACTAGTTCCACTGCCATCAAAACAAGGCAC----------
cfa-34c     CTGCTTA--------C--------------------------------------------
gga-34b     CGCCCCA--------CAATCACTAAATTCACTGCCATCAAAACAAGGCAC----------
cbr-34      AGG---TTGAC--AACGGCTACCTTCACTGCCACCCGAACATG-TAGTCCTC-------
cel-34      CC----TTGAC--AACGGCTACCTTCACTGCCACCCCGAACATG-TCGTCCATCTTTGAA
dme-34      AAT-TATTGCC--GTTGACAA--TTCACAGCCACTATCTTCAC--TGCCGCCGCGACAAG
dps-34      AAAATATTGCC--TTTGACCA--TTCACAGCCACTATCTTCAC--TGCCGCCGCGACAAG
bmo-34      AAA----TGAC--AACAGCCA---CTAACGACACTGCTCCTGCG-TGCACCCTAAATCA-
```

图 3 miR-34 基因家族前体的多序列比对

Fig. 3 Multiple sequence alignment of the precursor sequence of miR-34 gene family

星号表示同源性较高的碱基（Highly conserved bases are in asterisks）。

2.3 miR-34 系统发育分析

搜寻到的所有 miR-34 同源序列分别分布在节肢动物门、线虫纲、鱼纲、两栖纲、鸟纲和哺乳动物中。选择这些纲目中系统发育研究的模式物种用 MEGA 3.1 软件构建系统进化树（Jing et al，2004；Sempere et al，2006）（Fig. 4）。总的来看，进化树大致可分为三支：*mir-34b* 和 *mir-34c* 位于第一分支中；其次是 *mir-34a*；第三条分支中是 *mir-34*。从进化树的分支和距离上来看，*mir-34* 基因的分支出现在进化早期并且一直延续到现在，因而它很可能是 miR-34 基因家族中最早出现的基因形式，即祖先基因。*mir-34a* 基因在所有被建树研究的哺乳动物（人、黑猩猩、小鼠、大鼠、牛、狗）以及鸟类（原鸡），两栖纲动物（非洲爪蟾）中都保守，这一分支中只有一个例外是 *dre-mir-34*。正如前面基因位置分析结果那样，*mir-34b* 基因与 *mir-34c* 基因亲缘关系很近，并且它们与 *mir-34a* 基因是平行进化的。

在无脊椎动物中，miR-34 基因只有一个拷贝（*mir-34*），而在脊椎动物中 miR-34 基因有 3 个拷贝（*mir-34a*，*mir-34b*，*mir-34c*），且 miR-34 基因家族的 3 个成员稳定地存在于从鱼类到哺乳动物中。基于以上的鉴定和分析结果，可推测出 miR-34 基因家族可能的系统进化历史（Fig. 5）。*mir-34* 基因作为该家族最古老的基因形式仅存在于节肢动物和线虫纲动物等无脊椎动物中。在由无脊椎向脊椎

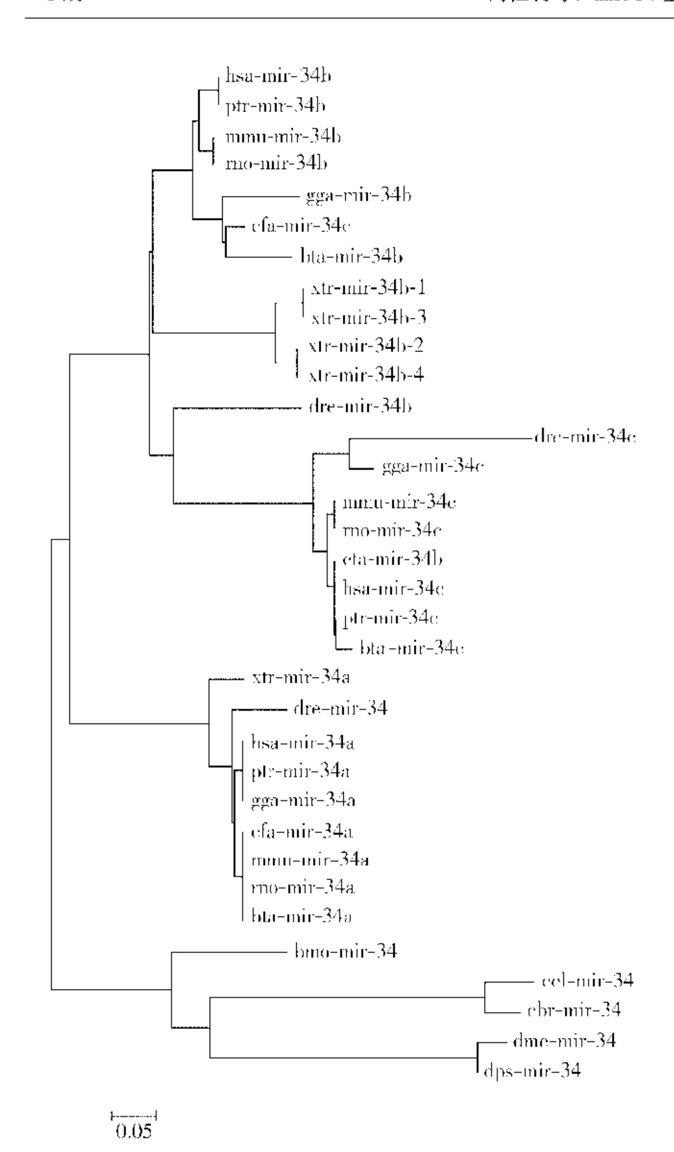


图 4　MEGA 软件 NJ 法构建 miR-34 基因家族系统进化树

Fig. 4　Phylogenetic tree of the miR-34 gene family constructed using MEGA version 3.1 based on the neighbor-joining method

缩写如下（Abbreviations are indicated as follows）：cel（线虫，*Caenorhabditis elegans*）；cbr（线虫，*C. briggsae*；dme（果蝇，*Drosophila melanogaster*）；dps（果蝇，*D. pseudoobscura*）；bmo（家蚕，*Bombyx mori*）；dre（斑马鱼，*Danio rerio*）；hsa（人，*Homo sapiens*）；ptr（黑猩猩，*Pan troglodytes*）；mmu（小鼠，*Mus musculus*）；rno（大鼠，*Rattus norvegicus*）；cfa（狗，*Canis familiaris*）；bta（牛，*Bos taurus*）；xtr（非洲爪蟾，*Xenopus tropicalis*）；gga（原鸡，*Gallus gallus*）。

动物进化的过程中新出现了紧密相连的 *mir-34b* 和 *mir-34c*，它们很可能是由 *mir-34* 基因经过一次局部重复，随后又发生了一次串联重复形成的。在进化到较高等的脊椎动物（两栖纲）过程中，*mir-34* 基因发生碱基变异后以 *mir-34a* 基因形式存在。此外，两栖纲动物中 *mir-34c* 的缺失可能是由于基因组序列拼接不完全的缘故。

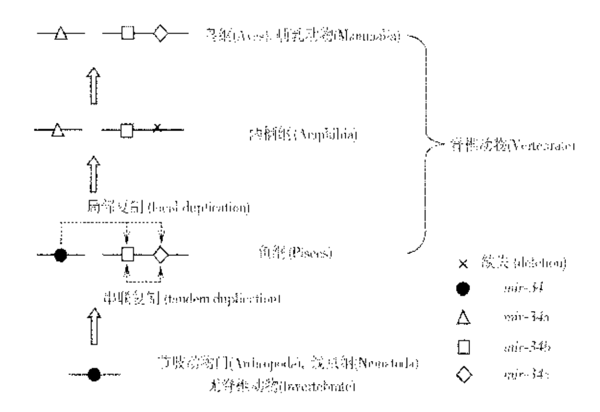


图 5　miR-34 基因家族可能的系统进化历史

Fig. 5　A plausible scenario for the evolution of the miR-34 gene family

3　讨　论

MicroRNA 作为生物体正常表达的内源非编码 RNA 分子，具有高度的保守性。Tanzer & Stadler（2004）在对 miR-17 基因簇的分子进化研究时发现，该基因簇进化可能与脊椎动物早期进化有关，进化过程伴随着个别基因的复制和缺失以及整个基因簇的复制。此外，水稻 miR-395 基因家族的分子进化研究结果也得出了相似的结论，基因家族的形成通常被认为是基因复制的结果（Elemento et al, 2002）。目前，在已测序的 33 种不同的动物中发现了 54 个 miR-34 基因。表明 miR-34 是高度保守的，广泛存在于后生动物中。在无脊椎动物和脊椎动物中 miR-34 基因家族成员个数并不相同，表明与 miR-17 基因簇和 miR-395 基因家族类似，miR-34 基因在从无脊椎动物向脊椎动物的进化过程中也发生了基因的串联重复和局部重复（Fig. 5）。在漫长的进化过程中，3 个成员组成的 miR-34 基因家族保守地存在于几乎所有的脊椎动物中，这是否意味着，与无脊椎动物相比，脊椎动物需要更多的 *mir-34* 基因来形成一个基因调控网络，从而成功的行使其调控功能；或者这些 *mir-34* 基因分别在脊椎动物的不同的组织中特异性表达。

同属非编码小 RNA 行列的还有 snoRNA 等。Zemann et al（2006）在研究线虫 snoRNA 进化时指出，snoRNA 基因的进化主要是通过反式重复（trans-duplication）和顺式重复（cis-duplication）的策略来完成的。Luo & Li（2006）在研究人的 box H/ACA snoRNA 时发现，snoRNA 还可以通过 RNA 介导的机制而复制。此外，snoRNA 与其 rRNA 靶序列存在共

进化的现象。非编码小 RNA 基因的进化有共通之处，miRNA 的进化也主要是通过重复及随后的序列变异来完成的（Pasquinelli & Ruvkun，2002；Grun et al，2005；Houbaviy et al，2005）。由于 miRNA 行使其功能是通过与其靶序列互补配对的方式而完成，它很可能也是与其靶基因共进化的（Allen et al，2004）。对非编码小 RNA 基因的系统进化分析将有助于进一步了解其起源和作用机制。

参考文献：

Allen E，Xie ZX，Gustafson AM，Sung GH，Spatafora JW，Carrington JC. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana* [J]. *Nat Gene*，**36**：1245-1246.

Altschul SF，Gish W，Miller W，Myers EW，Lipman DJ. 1990. Basic local alignment search tool [J]. *Mol Biol*，**215**（3）：403-410.

Ambros V，Bartel B，Bartel DP，Burge CB，Carrington JC，Chen X，Dreyfuss G，Eddy SR，Griffiths-Jones S，Marshall M，Matzke M，Ruvkun G，Tuschl T. 2003. A uniform system for microRNA annotation [J]. *RNA*，**9**（3）：277-279.

Doench JG，Sharp PA. 2004. Specificity of microRNA target selection in translational repression [J]. *Genes Dev*，**18**（5）：504-511.

Elemento O，Gascuel O，Lefranc MP. 2002. Reconstructing the duplication history of tandemly repeated genes [J]. *Mol Biol Evol*，**19**（3）：278-288.

Grun D，Wang YL，Langenberger D，Gunsalu KC，Rajewsky N. 2005. MicroRNA target predictions across seven Drosophila species and comparison to mammalian targets [J]. *PLoS Comput Biol*，**1**（1）：e13.

Houbaviy HB，Dennis L，Jaenisch R，Sharp PA. 2005. Characterization of a highly variable eutherian microRNA gene [J]. *RNA*，**11**（8）：1245-1257.

Houbaviy HB，Murray MF，Sharp PA. 2003. Embryonic stem cell-specific microRNAs [J]. *Dev Cell*，**5**（2）：351-358.

Jing ZZ，Cai XP. 2004. Progress of model organism genome [J]. *J Biomed Eng*，**21**（3）：506-511.［景志忠，才学鹏. 2004. 模式生物基因组研究进展. 生物医学工程杂志，**21**（3）：506-511.］

Kent WJ. 2002. Blat: The Blast-like alignment tool [J]. *Genome Res*，**12**（4）：656-664.

Kim VN，Nam JW. 2006. Genomics of microRNA [J]. *Genetics*，**22**（3）：165-173.

Lagos-Quintana M，Rauhut R，Lendeckel W，Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs [J]. *Science*，**294**（5543）：853-858.

Lau NC，Lim LP，Weinstein EG，Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans* [J]. *Science*，**294**：858-862.

Lee RC，Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans* [J]. *Science*，**294**：862-864.

Lee RC，Feinbaum RL，Ambros V. 1993. The *C. elegans* heterochronic gene *lin*-4 encodes small RNAs with antisense complementarity to *lin*-14 [J]. *Cell*，**75**（5）：843-854.

Llave C，Xie ZX，Kasschau KD，Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class off *Arabidopsis* miRNA [J]. *Science*，**297**（5589）：2053-2056.

Luo YP，Li SG. 2007. Genome-wide analyses of retorgenes derived from the human box H/ACA snoRNAs [J]. *Nucleic Acids Res*，**35**（2）：559-571.

Pasquinelli AE，Reinhart BJ，Slack F，Martindale MQ，Kuroda MI，Maller B，Hayward DC，Ball EE，Deqnan B，Muller P，Spring J，Srinivasan A，Fishman M，Finnerty J，Corbo J，Levine M，Leahy P，Davidson E，Ruvkun G. 2000. Conservation of the sequence and temporal expression of *let*-7 heterochronic regulatory RNA [J]. *Nature*，**408**：86-89.

Pasquinelli AE，Ruvkun G. 2002. Control of developmental timing by microRNAs and their targets [J]. *Annu Rev Cell Dev Biol*，**18**：495-513.

Reinhart BJ，Slack FJ，Basson M，Pasquinelli AE，Bettinger JC，Rougvie AE，Horvitz HR，Ruvkun G. 2000. The 21-nucleotide *let*-7 RNA regulates developmental timing in *Caenorhabditis elegans* [J]. *Nature*，**403**：901-906.

Reinhart BJ，Weinstein EG，Rhoades MW，Bartel B，Bartel DP. 2002. MicroRNAs in plants [J]. *Genes Dev*，**16**（13）：1616-1626.

Rodriguez A，Griffiths-Jones S，Ashurst JL，Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units [J]. *Genome Res*，**14**（10A）：1902-1910.

Ruvkun G. 2001. Molecular biology: Glimpses of a tiny RNA world [J]. *Science*，**294**（5543）：797-799.

Sempere LF，Cole CN，Mcpeek MA，Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint [J]. *J Exp Zoolog B Mol Dev Evol*，**306**（6）：575-588

Tanzer A，Stadler PF. 2004. Molecular evolution of a microRNA cluster [J]. *Mol Biol*，**339**（2）：327-335.

Thompson JD，Higgins DC，Gibson TJ. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice [J]. *Nucleic Acids Res*，**22**（22）：4673-4680.

Tong CZ，Jin YF，Zhang YZ. 2006. Computational prediction of microRNA genes in silkworm genome [J]. *Zhejiang Univ Sci B*，**7**（10）：806-816.

Zemann A，Bekke A O-D，Kiefmann M，Brosius J，Schmitz J. 2006. Evolution of small nucleolar RNAs in nematodes [J]. *Nucleic Acids Res*，**34**（9）：2676-2685.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction [J]. *Nucleic Acids Res*，**31**（13）：3406-3415.